05/12/2024

*Alessandro Allegranzi*

# ASSIGNMENT 1: DECISION TREES

The purpose of this assignment is to reinforce the idea of decision trees hands-on.

## Please use this Word file template for your response. Follow—and retain—these instructions in gray text. Insert your work in black where indicated. Keep in mind the evaluation matrix below as you do the work and use it to guide what you submit. Complete this assignment with the assistance of an AI generator such as ChatGPT. Your work will be assessed in terms of *your value added* as per the evaluation matrix at the end. You will describe your value added in the format provided by this Word template—it consists of your prompts together with you edits and additions to AI-generated material.

You may build on the application you select (see Section 1) for subsequent assignments if you wish.

Use no more than 6 pages of 12-point text excluding figures, the instructions in gray, your AI generator descriptions, and appendices. You can add as many appendices as you like, as per the instructions provided in the first appendix below.

You may build on the work of others but (1) show clearly that you understand this work and (2) observe all plagiarism rules scrupulously, including clear citations. Use the Reference section at the end.

## 1. SUMMARY DESCRIPTION

In one or two sentences, describe a decision tree application not in the literature, that you will implement. Select an application of personal or professional interest to you. You can begin with existing code (an example is [here](https://colab.research.google.com/drive/1DGZdokU6DOXCtDjYjGU_yev0_6KZOP2o?usp=sharing), which you can copy to your Google drive).

I will implement a decision tree application with a regression model that predicts the temperature and likelihood of rainfall in the city of Boston on a certain day based on historical weather data from the past 24 years. Weather forecasts are extremely useful, and a decision tree trained on historical data should be accurate enough for people to plan their day. A regression model will be used since the desired output is a numerical prediction and not a classification.

*To show I have run a decision tree for the draft here is a link to the notebook I am working on for this assignment:* [*https://colab.research.google.com/drive/13l4ouydDpBIzwLeN-oZWV7shWWL8OeJj?usp=sharing*](https://colab.research.google.com/drive/13l4ouydDpBIzwLeN-oZWV7shWWL8OeJj?usp=sharing)*. Apologize for the very rough state. Also below is a screenshot showing I trained a regression tree (this won’t be the actual finalized model I use, just to show I ran it):*

A screenshot of a computer program

Description automatically generated

A screenshot of a computer code

Description automatically generated

*^^ seems cold for July definitely need to better tune the training for the submission.*

### >>AI generation (or check: I did not use AI generation here \_X\_). Please collapse this section before submitting.

PARAGRAPH DESCRIBING YOUR VALUE ADDED TO AI-GENERATED MATERIAL

Your response replaces this.

YOUR PROMPT SEQUENCE

[1] Your first prompt replaces this.

[2]

Your response replaces this.

## 2. DATA SOURCE

Identify and explain the source of your data (e.g., collected by hand; Kaggle). Point to a URL so we can see the data.

I requested data from the National Centers for Environmental Information. (https://www.ncdc.noaa.gov/cdo-web/search). NOAA collects an extensive amount of information from weather stations around the US and is a reliable source of publicly available weather data. It’s not possible to actually visualize the data online, but requesting a download is easy. I will attach a CSV file with my submission to show the data gathered. I selected daily temperature and precipitation data from the Boston Logan weather station from January 1, 2000 to May 9, 2024.   
  
Sample:

A table with numbers and letters

Description automatically generated with medium confidence

## 3. I/O CASES

For three different inputs (e.g., a list or array) for your application, show the outputs.

**Input 1:** Date formatted day / month / year

**Output 1:** Expected average temp for the day

-------------------- -

**Input 2:** dictionary: date and temperature

**Output 2:** Expected precipitation for the day.

-------------------- -

**Input 3:** dictionary: date and temperature

**Output 3:** Expected snowfall for the day.

*Are three different i/o cases required? I’m having trouble thinking of additional inputs/outputs for the weather prediction tree. Was trying to keep assignment 1 scoped simply to start out. Should I shift to a more complex dataset?*

### >>AI generation (or check: I did not use AI generation here \_X\_). Please collapse this section before submitting.

## 4. KEY SOURCE CODE

Please supply the most relevant decision tree code).

### >>AI generation (or check: I did not use AI generation here \_\_). Please collapse this section before submitting.

PARAGRAPH DESCRIBING YOUR VALUE ADDED TO AI-GENERATED MATERIAL

Your response replaces this.

YOUR PROMPT SEQUENCE

[1] Your first prompt replaces this.

[2]

Your response replaces this.

Your response replaces this.

PARAGRAPH DESCRIBING YOUR VALUE ADDED TO AI-GENERATED MATERIAL

Your response replaces this.

YOUR PROMPT SEQUENCE

[1] Your first prompt replaces this.

[2]

Your response replaces this.

## 5. YOUR SOURCE

Please supply the URL of your code (e.g., as shared Colab code).

Colab link: https://colab.research.google.com/drive/13l4ouydDpBIzwLeN-oZWV7shWWL8OeJj?usp=sharing

### >>AI generation (or check: I did not use AI generation here \_\_). Please collapse this section before submitting.

PARAGRAPH DESCRIBING YOUR VALUE ADDED TO AI-GENERATED MATERIAL

Your response replaces this.

YOUR PROMPT SEQUENCE

[1] Your first prompt replaces this.

[2]

Your response replaces this.

Your response replaces this.

## 6. DATA ALTERATION FOR CHANGED RESULTS

Alter the data from your data source so that it changes in the outputs for the above inputs—but not as in Section 7 below. Explain in the format below why the output changes the way it does. Limit (for this Section 6): 2 normal paragraphs (remember that you can use appendices for reference material).

## 6.1 What Data Was Altered?

Your response replaces this.

## 6.2 How Did the Output Change?

Your response replaces this.

## 6.3 Why Did the Output Change in This Way?

Your response replaces this.

### >>AI generation (or check: I did not use AI generation here \_\_). Please collapse this section before submitting.

PARAGRAPH DESCRIBING YOUR VALUE ADDED TO AI-GENERATED MATERIAL

Your response replaces this.

YOUR PROMPT SEQUENCE

[1] Your first prompt replaces this.

[2]

Your response replaces this.

## 7. INCONSISTENT DATA

Alter the data so that it contains *inconsistencies*. Using the format below, show changes in the I/O of the examples in part 3, and explain. Limit: 3 normal paragraphs.

## 7.1 What Data Was Altered to be Inconsistent?

Your response replaces this.

## 7.2 How Did the Output Change?

Your response replaces this.

## 7.3 Why Did the Output Change in This Way?

Your response replaces this.

### >>AI generation (or check: I did not use AI generation here \_\_). Please collapse this section before submitting.

PARAGRAPH DESCRIBING YOUR VALUE ADDED TO AI-GENERATED MATERIAL

Your response replaces this.

YOUR PROMPT SEQUENCE

[1] Your first prompt replaces this.

[2]

Your response replaces this.

## 8. BENEFITS

In at most 3/4 page (of 12-point text), explain the pros and cons of decision trees *applied to the application you have chosen* (i.e., don’t respond generically).

**Pro’s of decision tree usage for my application**

Your response replaces this.

**Con’s of decision tree usage for my application**

Your response replaces this.

### >>AI generation (or check: I did not use AI generation here \_\_). Please collapse this section before submitting.

PARAGRAPH DESCRIBING YOUR VALUE ADDED TO AI-GENERATED MATERIAL

Your response replaces this.

YOUR PROMPT SEQUENCE

[1] Your first prompt replaces this.

[2]

Your response replaces this.

## 9. EVALUATION



# References

If you use anyone’s work, the extent of that usage must be made very clear. Where this use is via an AI generator, please use the sections provided. Every other reference must be cited at least once within the paper above.

[1] …

[2] …

# APPENDIX 1 (if required)

…

# APPENDIX 2 (if required)

…